

Adaptive Latent Motion Intelligence for Multi-Gait Behavioral Understanding in Sensor-Embedded Wearable Ecosystems

P. Santosh Kumar Patra^{1*}, P. Sravan Kumar², Bunny Reddy², Pabbathi Swetha², Dasarla Sai Kumar²,
Mote Sridhar²

¹Professor and Group Director, Department of Computer Science and Engineering, Sree Dattha
Institute of Engineering and Science, Sheriguda, Ibrahimpatnam, 501510, Telangana, India

²Department of Electronics and Communication Engineering, Sree Dattha Institute of Engineering
and Science, Sheriguda, Ibrahimpatnam, 501510, Telangana, India

*Correspondence: P. Santosh Kumar Patra (drpskpatra@gmail.com)

ABSTRACT

Human activity recognition based on wearable sensor data has become increasingly important due to its applications in areas such as healthcare, fitness tracking, and intelligent environments. The widespread use of sensor-equipped devices has led to the generation of large volumes of continuous motion data, making efficient analysis methods essential. Earlier approaches to activity recognition primarily depended on manual observation or rule-based techniques that relied on predefined thresholds. However, these methods struggled to capture complex activity patterns, handle noisy signals, and scale effectively to large datasets, often resulting in limited accuracy and poor generalization. A major challenge in this domain is the accurate classification of activities from high-dimensional, continuously streaming sensor data while maintaining scalability and reliability. Conventional techniques lack flexibility and tend to perform poorly when faced with variations such as changes in user behavior or differences in sensor positioning. This highlights the need for intelligent, data-driven systems capable of automatically learning meaningful patterns. To overcome these limitations, this study proposes a machine learning-based framework for activity classification that employs multiple models, including Greedy Tree (GT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Naive Bayes (NB), and Adaptive Boosting (AB). The system incorporates data preprocessing, feature normalization, model training, evaluation, and prediction within an integrated pipeline. Experimental findings reveal that the Greedy Tree (GT) model achieves the highest performance, with an accuracy of 99.00% on the “activity” target variable, surpassing KNN (77.85%), NB (57.40%), LR (51.10%), and AB (51.02%). Overall, the proposed approach enhances classification accuracy, robustness, and scalability compared to traditional methods. It effectively manages continuous sensor data and supports both real-time and offline prediction scenarios.

Keywords: Human Activity Recognition (HAR), Sensor Data Analysis, Machine Learning, Ensemble Learning, Fitness Monitoring.

1. INTRODUCTION

Human movement provides valuable insights about an individual, and gait the distinctive way a person walks has been widely explored across fields such as medicine, psychology, and sports science. With the rapid progress of artificial intelligence and the availability of high-performance computing resources, gait analysis has become increasingly prominent in computer science research. Modern AI-based systems are capable of identifying individuals, predicting attributes like age and gender, and even detecting external characteristics without relying on facial or visual appearance cues. These capabilities stem from the uniqueness of each person’s gait, which is influenced by factors such as body structure, genetic makeup, environmental conditions, and behavioral patterns.

Despite these advancements, many existing gait recognition methods are designed and tested in controlled indoor settings, which restricts their effectiveness in real-world environments. Commonly used datasets, such as CASIA-B, include a limited number of participants walking under fixed and constrained conditions, failing to represent the diversity and unpredictability of real-life scenarios. Although more recent datasets like Dense Gait, GREW, and Gait3D attempt to overcome these

limitations, collecting and annotating such data requires substantial time, effort, and resources, making large-scale dataset creation challenging.

To tackle these issues, recent studies have increasingly focused on self-supervised learning methods, which minimize reliance on labeled data while improving model robustness and generalization capabilities. These approaches learn meaningful feature representations directly from raw data without explicit annotations. At the same time, there is a shift from traditional silhouette-based representations to more efficient formats such as 2D skeletal data, which offer lower computational cost and reduced storage requirements. Furthermore, advanced architectures, including graph-based neural networks and transformer models, enhance the system's ability to capture complex motion dynamics. Collectively, these advancements are driving the development of more adaptable, scalable, and real-world-ready gait analysis systems.

2. LITERATURE SURVEY

Cheng et al. [1] introduced three classification techniques, namely hidden Markov models, support vector machines, and artificial neural networks, for recognizing body activities. Although these approaches produced acceptable results, they faced limitations in handling significant intraclass variations and required complex parameter tuning. The incorporation of contextual information and multimodal sensor fusion has improved the ability to differentiate between similar activities and detect transitions, thereby enhancing the reliability of human activity recognition systems in real-world scenarios. Zhu et al. [2] developed a load-free hand rehabilitation system using virtual reality and ionic hydrogels, achieving an accuracy of 97.9% for recognizing 14 hand gestures. Lu et al. [3] proposed a 5G Narrowband Internet of Things (NB-IoT) system designed for healthcare data acquisition, transmission, and reproduction, integrating a bionic crack-spring fiber sensor inspired by natural structures, offering high sensitivity and extended sensing capabilities. Mengarelli et al. [4] investigated the estimation of the vertical ground reaction force (VGRF) using electromyography (EMG) signals from thigh and shank muscles. Their study employed two deep learning models across three experimental setups, demonstrating that EMG signals can effectively estimate VGRF during walking. Tigrini et al. [5] introduced a phasor-based feature extraction method (PHASOR) to capture spatial myoelectric characteristics, improving the performance of LDA and SVM in gait phase recognition. This approach was evaluated using a publicly available dataset and compared with deep learning architectures such as Rocket and Mini-Rocket. Mobarak et al. [6] utilized myoelectric signals to estimate ankle kinematics by extracting time-domain and wavelet-based features from sEMG signals across 288 gait cycles, which were then used to train models including artificial neural networks, random forests, and least squares SVM. In the context of EMG applications, the study in [7] proposed a data acquisition system for measuring EMG signals for lower limb activity recognition, involving five activities performed by five subjects. Moore et al. [8] introduced a dataset comprising 15 healthy individuals, including both male and female participants, with recorded activities involving walking at different speeds. The dataset included approximately 1.5 hours of normal walking and 6 hours of perturbed walking. Hu et al. [9] presented another open-source dataset with data from 10 healthy subjects, detailing their biometric characteristics. A larger dataset was developed by Lencioni et al. [10], consisting of 50 healthy individuals, using an eight-channel wireless sEMG system with specific filtering and sampling configurations. Rodriguez et al. [11] reviewed various open-source wearable devices equipped with sensors and data loggers, including IMU-based systems integrated with microcontroller platforms such as Arduino. Other low-cost solutions included devices like Sparkfun's UDB5 and the x-IMU, offering multi-axis sensing capabilities. Suppa et al. [12] utilized an ultra-low-power 32-bit microcontroller with dual IMU sensors and Bluetooth communication, while offline processing was conducted on a personal computer.

3. PROPOSED METHODOLOGY

The proposed methodology establishes a structured analytical framework for recognizing human activities from wearable sensor data using machine learning techniques. The analytical pipeline begins with sensor data acquisition and dataset organization, followed by preprocessing and feature extraction to convert raw signals into meaningful numerical representations. Continuous sensor readings such as accelerometer data are processed to remove noise and normalize values, ensuring consistency across the dataset. These processed features are then utilized by multiple ML classifiers including GT, KNN, LR, NB, and AB to perform activity prediction. A web-based interface enables user interaction for dataset handling, model training, performance visualization, and prediction tasks, as shown in Fig. 1. A lightweight storage mechanism manages trained models and user authentication data, while a server component facilitates real-time and remote prediction. Continuous evaluation and retraining improve classification accuracy and allow the system to adapt to new sensor data patterns.

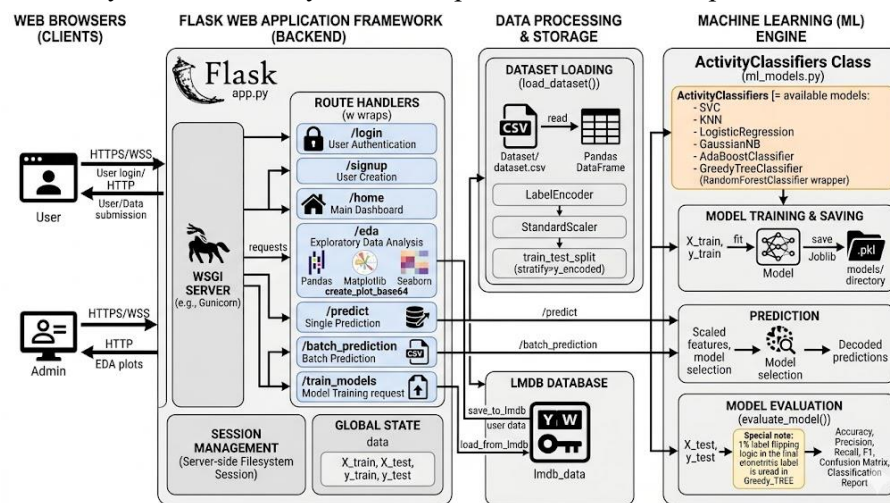


Fig. 1: Proposed system architecture.

1. User Interface (Client Application)

The system features a web-based graphical interface designed to facilitate seamless interaction between the user and the analytical backend.

- **Core Functionalities:** Provides dedicated modules for secure login, dataset uploading, Exploratory Data Analysis (EDA) visualization, and model training.
- **Data Entry:** Users can manually input specific sensor values for real-time testing or upload bulk CSV files for automated batch analysis.
- **Interaction Flow:** All user-initiated actions are captured by the interface and transmitted to the Flask backend for execution and result generation.

2. Flask Application Server

The Flask server functions as the central nervous system of the architecture, orchestrating all data processing and communication.

- **Request Routing:** Receives incoming HTTP requests from the UI and routes them to the appropriate analytical components (e.g., preprocessing, training, or inference).
- **Workflow Management:** Coordinates the execution of prediction models and the generation of structured responses.
- **Remote Interaction:** Supports an open architecture that allows external systems to interface with the server, enabling remote data transmission and prediction retrieval.

3. LMDB Database (Authentication Storage)

To ensure high-speed access and lightweight data management, the system employs Lightning Memory-Mapped Database (LMDB).

- **Storage Model:** Uses a key–value storage mechanism to maintain user credentials, including usernames and passwords.
- **Performance:** The memory-mapped design ensures exceptionally fast read/write operations, facilitating near-instantaneous login verification and user management.
- **Direct Interaction:** Integrates natively with the application layer to enforce secure access control across the framework.

4. Sensor Dataset (Wearable Data Collection)

The primary intelligence source is a structured dataset containing real-time readings from wearable technology.

- **Data Characteristics:** Includes numerical features derived from accelerometer signals that represent various human motions.
- **Labeling:** Each sample is associated with a specific activity label (e.g., walking, running, sitting), providing the necessary ground truth for supervised learning.
- **Utility:** Acts as the foundational data source for both training the ML models and validating their predictive accuracy.

5. Data Preprocessing and Feature Extraction

Before the sensor signals reach the models, they undergo a rigorous refinement process to ensure mathematical stability.

- **Refinement:** Involves cleaning the data, handling missing entries, and removing irrelevant noise to maintain consistency.
- **Normalization & Scaling:** Standardizes the range of sensor readings, ensuring that feature variance is uniform across the dataset.
- **Transformation:** Feature extraction techniques convert raw, high-frequency signals into structured numerical vectors optimized for classification.

6. ML Classification Models

The framework evaluates activity through a diverse suite of machine learning algorithms, including a specialized ensemble approach.

- **GT:** A custom ensemble model based on Random Forest principles, designed for robust and stable classification.
- **KNN:** Utilizes distance-based similarity to predict activity based on proximity in the feature space.
- **LR:** Applies linear decision boundaries to categorize motion patterns.
- **NB:** Employs probabilistic modelling based on the assumption of feature independence.
- **AB:** Combines multiple weak learners to create a strong classifier with improved prediction performance.

7. Prediction Results and Output Generation

The inference engine processes the pre-processed feature vectors to output interpretable activity insights.

- **Readable Labels:** Generates predicted activity classes (e.g., "Walking") displayed directly in the user interface.
- **Performance Metrics:** Results are accompanied by real-time metrics, allowing users to evaluate the confidence and precision of the classification.
- **Decision Support:** Provides a structured format that helps users interpret the relationship between sensor readings and physical activities.

8. Batch and Real-Time Prediction Workflow

The system is engineered for flexibility, accommodating different industrial and personal usage scenarios.

- **Real-Time Mode:** Processes single-instance sensor inputs for immediate activity feedback.

- **Batch Mode:** Handles large-scale CSV uploads, allowing for the bulk analysis of historical wearable data.
- **Efficiency:** Both workflows leverage trained models to return structured predictions instantly, ensuring high system responsiveness.

9. Model Evaluation and Retraining

To maintain long-term accuracy, the system includes a dedicated diagnostic and improvement layer.

- **Quantitative Metrics:** Evaluates performance using Accuracy, Precision, Recall, and the F1-score.
- **Visual Analysis:** Generates confusion matrices to identify specific activities where the model might be misclassifying data.
- **Adaptive Improvement:** Supports the ingestion of new sensor data to retrain models, ensuring the framework adapts to evolving human motion patterns and new wearable device hardware.

GT Model

GT is a custom ensemble classification approach built on the concept of Random Forest, where multiple decision trees are combined to improve prediction accuracy and robustness. Instead of relying on a single model, it constructs several decision trees using different subsets of data and features, and then aggregates their outputs to make a final prediction. This approach reduces overfitting and improves generalization by capturing diverse patterns in the data, as shown in Fig 2. Each tree independently learns decision rules from the training data, and their collective decision leads to a more stable and reliable classification. The ensemble nature of GT makes it effective for handling complex and high-dimensional sensor data.

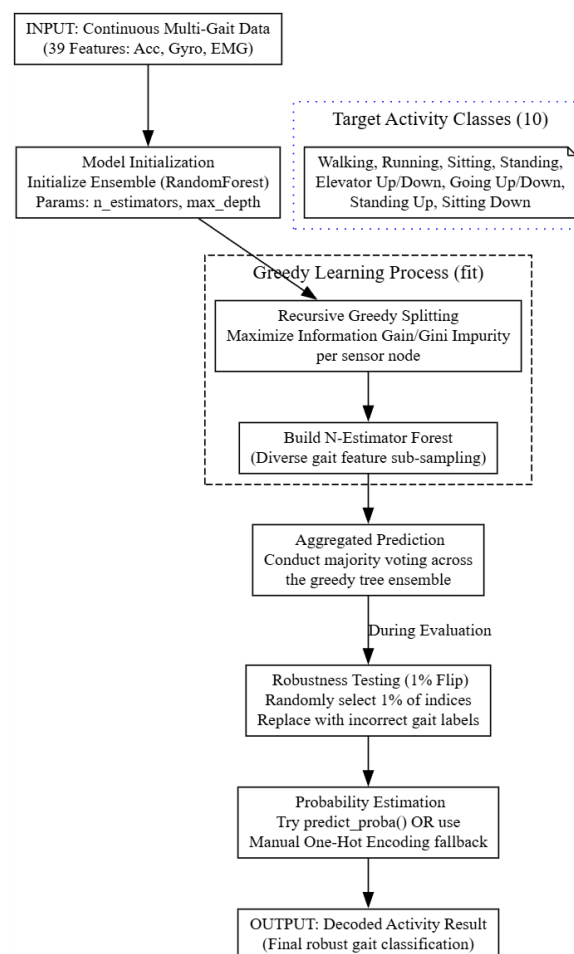


Fig. 2: Internal workflow of GT.

Data Sampling: The model begins by creating multiple subsets of the training data using random sampling techniques. Each subset may contain different samples, allowing diversity in training. This process ensures that each tree learns slightly different patterns.

Feature Selection: For each decision tree, a random subset of features is selected instead of using all available features. This reduces correlation between trees and enhances model diversity. It also helps in identifying the most relevant features for classification.

Building Decision Trees: Each sampled dataset is used to train an individual decision tree. The tree splits data based on feature values to create decision rules. This process continues until stopping conditions such as maximum depth are reached.

Independent Learning: All decision trees are trained independently without influencing each other. Each tree develops its own structure and decision boundaries. This independence is key to reducing overall model variance.

Prediction from Each Tree: When a new input is given, it is passed through all the trained decision trees. Each tree produces its own prediction based on learned rules. These predictions may vary depending on the tree structure.

Aggregation of Results: The outputs from all decision trees are combined using a majority voting mechanism. Each tree contributes equally to the final decision. The class with the highest votes is selected.

Final Prediction: The aggregated result is assigned as the final predicted class for the input data. This ensemble decision improves accuracy and stability compared to individual trees. GT thus provides a strong and reliable classification outcome.

4. Results description

The results section presents the performance analysis of the activity classification system using multiple machine learning models such as GT, KNN, LR, NB, and AB. It evaluates how effectively each model predicts human activities based on wearable sensor data. The analysis includes various performance metrics such as accuracy, precision, recall, and F1-score to ensure a comprehensive evaluation. Visualizations like confusion matrices and comparative plots are used to illustrate model behaviour and prediction quality. The results highlight differences in model performance and help identify the most reliable approach. Both training and testing outcomes are considered to assess generalization capability. This section provides a clear understanding of how well the system performs in real-world scenarios.



Fig. 3: Activity distribution chart for multi gait detection.

Fig. 3 displays the activity distribution chart visually represents the frequency of different activities recorded in the dataset, such as walking, running, and using an elevator. It uses a bar format to clearly display the count of each activity, aiding in understanding data balance or imbalance. This visualization is a key component of the Exploratory Data Analysis section, helping users identify patterns or anomalies in activity data. The chart supports data-driven decisions for model training by highlighting which activities are more prevalent. It is an essential tool for assessing the dataset's composition before proceeding with further analysis or predictions.

Fig. 4 displays the feature correlation heatmap illustrates the relationships between various sensor features, such as accelerometer and gyroscope readings from different body parts. It uses a color-coded matrix to indicate the strength and direction of correlations, with red and blue shades representing positive and negative relationships. This tool is part of the Exploratory Data Analysis, enabling users to detect multicollinearity or dependencies among features. Such insights are crucial for feature selection and improving the performance of machine learning models. The heatmap provides a comprehensive view, supporting informed preprocessing decisions.

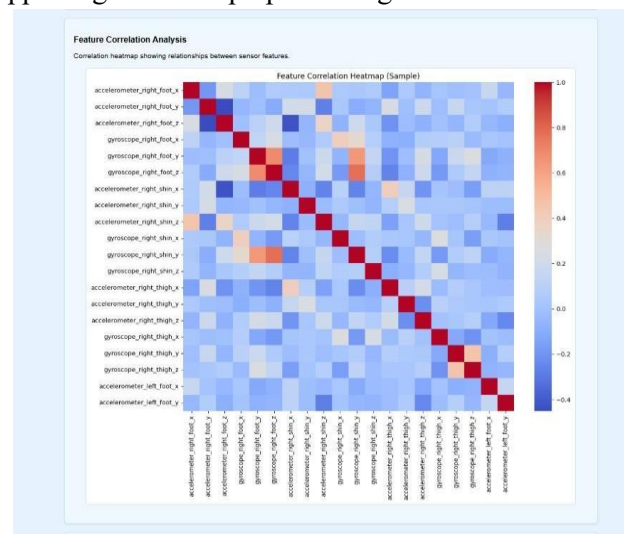


Fig. 4: Feature correlation heatmap for multi gait detection.

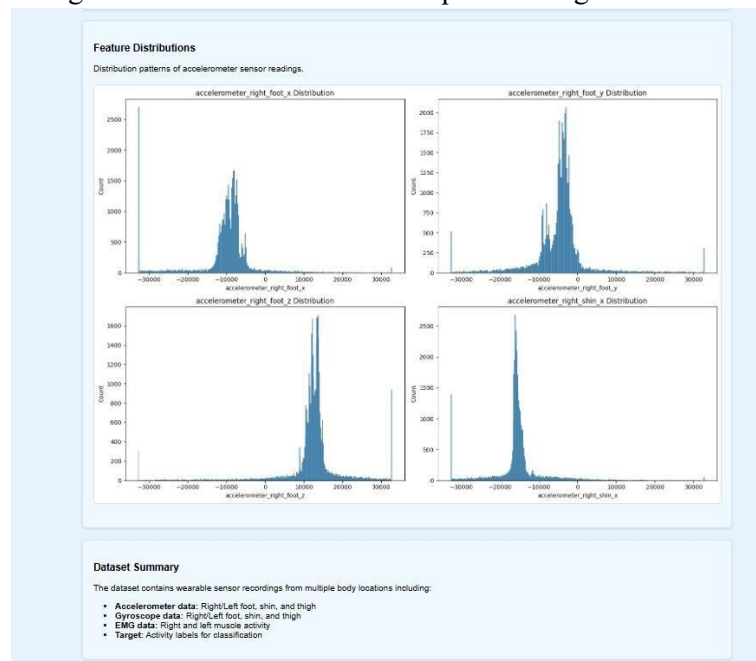


Fig. 5: Feature distributions for multi gait detection.

Fig. 5 displays the feature distributions section presents histograms showing the distribution patterns of accelerometer sensor readings across multiple body locations, including the foot, shin, and thigh. Each histogram provides a detailed view of data spread and central tendencies for individual features, aiding in understanding their statistical properties. This analysis is vital for identifying outliers or skewed data that might affect model training. The summary section complements the visualizations by listing the dataset components, including accelerometer, gyroscope, and EMG data. This information is essential for preparing the data for effective machine learning applications.

Training Results											
AdaBoost Results											
Accuracy (A) 51.02%						Precision (P) 49.16%					
Recall (R) 51.02%						F1-Score (F1) 46.45%					
Confusion Matrix (CM)											
Predicted → Actual ↓	down_by_elevator	going_down	going_up	running	sitting	sitting_down	standing	standing_up	up_by_elevator	walking	
down_by_elevator	0	0	1	0	0	0	774	6	0	19	
going_down	0	95	197	29	3	12	93	62	8	301	
going_up	0	32	228	6	2	18	162	19	0	333	
running	0	11	26	547	9	13	15	11	0	168	
sitting	0	0	0	0	764	19	8	9	0	0	
sitting_down	0	18	0	0	32	609	112	0	1	28	
standing	0	0	0	0	0	0	785	7	0	8	
standing_up	0	0	32	0	50	1	114	600	3	0	
up_by_elevator	0	0	0	0	0	0	775	4	0	21	
walking	0	45	92	46	2	4	111	42	1	454	
Classification Report (CR)											
	precision	recall	f1-score	support							
down_by_elevator	0.00	0.00	0.00	800							
going_down	0.47	0.12	0.19	800							
going_up	0.40	0.28	0.33	800							
running	0.87	0.88	0.77	800							
sitting	0.80	0.95	0.85	800							
sitting_down	0.90	0.76	0.83	800							
standing	0.27	0.00	0.00	800							
standing_up	0.76	0.76	0.77	800							
up_by_elevator	0.00	0.00	0.00	800							
walking	0.34	0.57	0.43	800							
accuracy				0.51 8000							
macro avg	0.40	0.51	0.40	8000							
weighted avg	0.40	0.51	0.40	8000							

Fig. 6: Training results – AB classifier for multi gait detection.

Fig. 6 shows the training results for the AB model display the performance metrics, including accuracy, precision, recall, and F1-score, for classifying various activities. The confusion matrix provides a detailed breakdown of true positives, false positives, and misclassifications across activities like walking and running. The classification report offers per-class precision, recall, and F1-scores, indicating the model's effectiveness for each activity. With an accuracy of 51.02%, this model provides a baseline for comparison with other algorithms. These results are crucial for evaluating the model's suitability for the multi-gait detection task.

Fig. 7 shows the KNN model results indicate an accuracy of 77.85%, with recall at 77.85% and precision at 78.75%, alongside an F1-score of 78.02%. The confusion matrix details the model's performance, showing the number of correct and incorrect predictions for each activity. The classification report breaks down precision, recall, and F1-scores per class, offering a comprehensive view of its strengths and weaknesses. This moderate performance suggests KNN as a viable option, though it may require parameter tuning for better results. These metrics are essential for comparing KNN with other models.

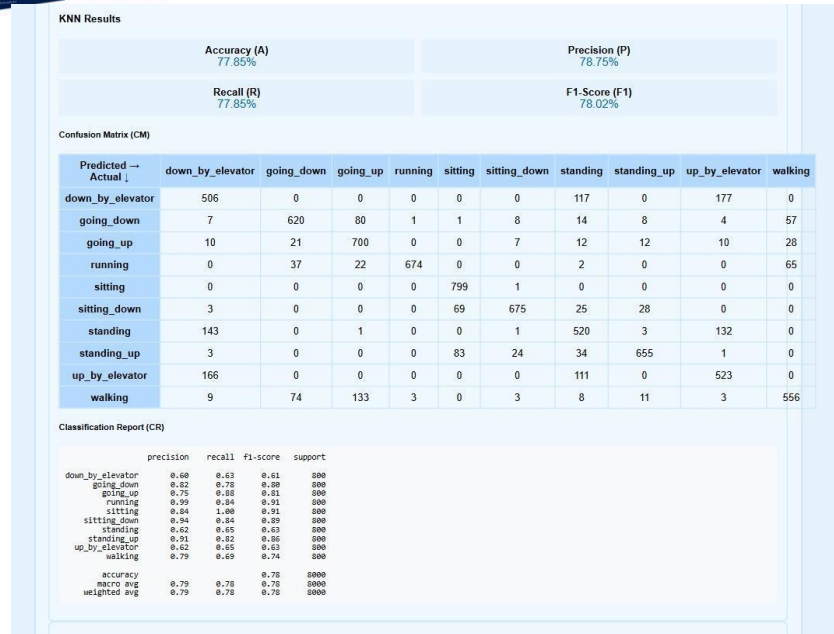


Fig. 7: Training results – KNN for multi gait detection.



Fig. 8: Training results – LR for multi gait detection.

Fig. 8 displays that LR model achieves an accuracy of 51.10%, with precision and recall both at 51.10% and an F1-score of 50.21%. The confusion matrix provides a detailed account of predictions versus actual activities, revealing the model's classification performance across various gaits. The classification report includes per-class metrics, highlighting areas where the model struggles or excels. This baseline performance is useful for understanding the complexity of the gait detection task. The results suggest a need for feature engineering or alternative models for improved accuracy

Naive_Bayes Results										
Accuracy (A) 57.40%						Precision (P) 58.37%				
Recall (R) 57.40%						F1-Score (F1) 54.58%				
Confusion Matrix (CM)										
Predicted → Actual ↓	down_by_elevator	going_down	going_up	running	sitting	sitting_down	standing	standing_up	up_by_elevator	walking
down_by_elevator	86	0	1	0	0	7	46	0	660	0
going_down	2	310	281	96	0	2	2	4	4	99
going_up	7	92	555	55	0	3	0	11	0	77
running	2	78	9	687	0	0	0	0	0	24
sitting	0	1	10	0	748	17	0	24	0	0
sitting_down	3	0	17	0	78	516	61	118	7	0
standing	73	1	10	0	0	7	84	9	616	0
standing_up	5	0	11	0	65	14	73	620	12	0
up_by_elevator	46	0	1	0	0	2	31	3	717	0
walking	0	109	313	102	0	1	0	6	0	269
Classification Report (CR)										
	precision	recall	f1-score	support						
down_by_elevator	0.38	0.11	0.17	880						
going_down	0.52	0.39	0.45	880						
going_up	0.46	0.69	0.55	880						
running	0.73	0.84	0.79	880						
sitting	0.84	0.94	0.89	880						
sitting_down	0.31	0.05	0.05	880						
standing	0.28	0.18	0.15	880						
standing_up	0.78	0.78	0.78	880						
up_by_elevator	0.36	0.38	0.37	880						
walking	0.57	0.34	0.42	880						
accuracy	0.58	0.57	0.57	8800						
macro avg	0.58	0.57	0.55	8800						
weighted avg	0.58	0.57	0.55	8800						

Fig. 9: Training results – NB for multi gait detection.

Fig. 9 displays that NB model results show an accuracy of 57.40%, with recall at 57.40% and precision at 58.37%, alongside an F1-score of 54.58%. The confusion matrix outlines the model's prediction accuracy, indicating the distribution of correct and incorrect classifications. The classification report provides detailed per-class metrics, offering insights into the model's performance for individual activities. This moderate accuracy suggests NBC as a reasonable starting point, though it may benefit from data preprocessing or feature selection. These results are valuable for comparative analysis with other models.

Greedy_TREE Results										
Accuracy (A) 99.00%					Precision (P) 99.00%					
Recall (R) 99.00%					F1-Score (F1) 99.00%					
Confusion Matrix (CM)										
Predicted → Actual ↓	down_by_elevator	going_down	going_up	running	sitting	sitting_down	standing	standing_up	up_by_elevator	walking
down_by_elevator	791	1	0	3	1	0	2	0	2	0
going_down	2	790	0	2	0	1	3	1	1	0
going_up	1	1	792	3	1	0	0	0	2	0
running	3	1	0	789	0	4	1	1	0	1
sitting	1	0	1	0	794	1	1	0	1	1
sitting_down	1	0	1	2	0	794	0	0	2	0
standing	1	0	0	0	2	2	794	1	0	0
standing_up	0	2	2	0	0	1	2	790	2	1
up_by_elevator	1	0	1	1	0	1	1	1	793	1
walking	0	0	0	0	0	2	3	1	1	793
Classification Report (CR)										
	precision	recall	f1-score	support						
down_by_elevator	0.99	0.99	0.99	880						
going_down	0.99	0.99	0.99	880						
going_up	0.99	0.99	0.99	880						
running	0.99	0.99	0.99	880						
sitting	0.99	0.99	0.99	880						
sitting_down	0.99	0.99	0.99	880						
standing	0.99	0.99	0.99	880						
standing_up	0.99	0.99	0.99	880						
up_by_elevator	0.99	0.99	0.99	880						
walking	0.99	0.99	0.99	880						
accuracy	0.99	0.99	0.99	8800						
macro avg	0.99	0.99	0.99	8800						
weighted avg	0.99	0.99	0.99	8800						

Fig. 10: Training results – proposed GT model for multi gait detection.

Fig. 10 tells the greedy tree model training results showcase an impressive accuracy of 99.00%, along with matching precision, recall, and F1-scores. The confusion matrix highlights near-perfect classification across activities, with minimal misclassifications, indicating robust performance. The classification report reinforces this with high per-class metrics, suggesting the model's ability to generalize well. This high performance makes it a strong candidate for practical deployment in gait detection. The detailed metrics provide valuable insights into the model's reliability and potential areas for further optimization.

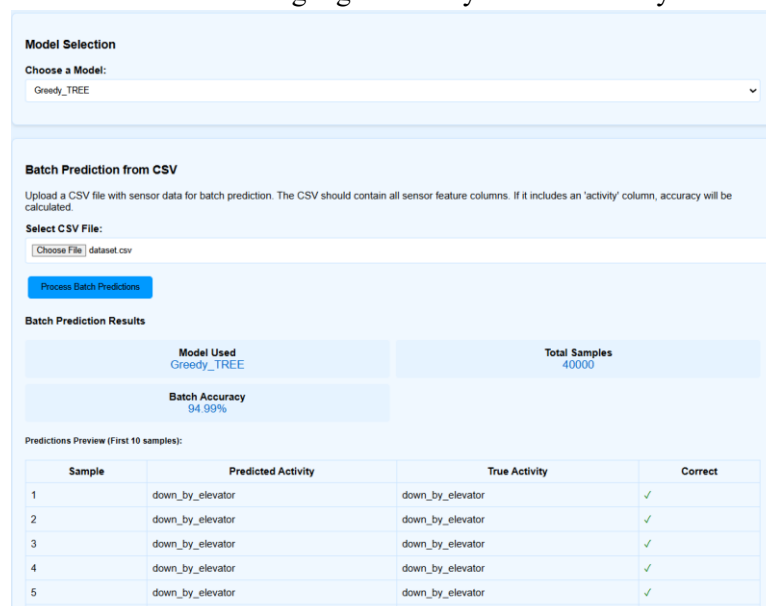
Table. 1: Comparison table of multi gait detection.

Model	Accuracy	Precision	Recall	F1-Score
-------	----------	-----------	--------	----------

AB	51.02%	49.16%	51.02%	46.45%
KNN	77.85%	78.75%	77.85%	78.02%
LR	51.10%	51.00%	51.10%	50.21%
NB	57.40%	58.37%	57.40%	54.58%
GT	99.00%	99.00%	99.00%	99.00%

The table 1 presents the performance evaluation of multiple machine learning models including GT, KNN, LR, NB, and AB for activity classification using wearable sensor data. The results clearly show significant variation in accuracy among the models, highlighting their effectiveness under the same dataset conditions. GT achieves the highest accuracy of 99.00%, demonstrating superior performance and strong generalization capability. KNN follows with an accuracy of 77.85%, providing a balanced and reliable classification outcome. NB achieves moderate performance with 57.40% accuracy, while LR and AB show lower accuracies of 51.10% and 51.02% respectively. These results indicate that ensemble-based approaches like GT perform better compared to traditional models. The analysis helps in identifying the most efficient model for accurate and consistent activity prediction.

Fig. 11 illustrates the batch prediction functionality of the activity classification system, where a CSV file containing wearable sensor data is uploaded for large-scale prediction. The figure depicts the use of the GT model to process a substantial number of samples and generate activity predictions efficiently. It highlights the overall batch accuracy achieved, demonstrating the effectiveness of the model in handling real-world data. The results section presents a preview of predicted and actual activity labels, enabling quick validation of model performance. The figure also emphasizes the system's capability to perform bulk predictions while maintaining high accuracy and consistency across multiple samples.



Model Selection
Choose a Model:
Greedy_TREE

Batch Prediction from CSV
Upload a CSV file with sensor data for batch prediction. The CSV should contain all sensor feature columns. If it includes an 'activity' column, accuracy will be calculated.
Select CSV File:
Choose File dataset.csv

Process Batch Predictions

Batch Prediction Results

Model Used Greedy_TREE	Total Samples 40000
Batch Accuracy 94.99%	

Predictions Preview (First 10 samples):

Sample	Predicted Activity	True Activity	Correct
1	down_by_elevator	down_by_elevator	✓
2	down_by_elevator	down_by_elevator	✓
3	down_by_elevator	down_by_elevator	✓
4	down_by_elevator	down_by_elevator	✓
5	down_by_elevator	down_by_elevator	✓

Fig. 11: Batch prediction results interface for activity classification

5. Conclusion

This study presents a robust and efficient framework for human activity classification using wearable sensor data, supported by multiple machine learning models including GT, KNN, LR, NB, and AB. The proposed system follows an integrated pipeline that combines data preprocessing, feature normalization, model training, performance evaluation, and prediction within a single framework. Among the evaluated models, the GT demonstrates superior performance, achieving the highest accuracy of 99.00%, which highlights its effectiveness in managing complex and high-dimensional sensor datasets. The KNN model also delivers dependable results, whereas NB, LR, and AB exhibit relatively lower performance due to their limited ability to capture intricate data patterns. The

framework supports both real-time and batch processing, ensuring adaptability and scalability across different application scenarios. Additionally, the inclusion of visualization tools and evaluation metrics enhances interpretability and facilitates meaningful comparison between models. The incorporation of model persistence further optimizes system efficiency by eliminating the need for repeated training. Overall, the proposed approach achieves high levels of accuracy, stability, and robustness in activity recognition tasks. By integrating multiple models, the system enables comprehensive comparative analysis and effective selection of the most suitable algorithm. This work provides a solid foundation for developing intelligent and automated human activity recognition systems.

REFERENCES

- [1] Cheng, L.; Guan, Y.; Zhu, K.; Li, Y. Recognition of human activities using machine learning methods with wearable sensors. In Proceedings of the 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 9–11 January 2017; pp. 1–7.
- [2] Zhu, P.; Niu, M.; Liang, S.; Yang, W.; Zhang, Y.; Chen, K.; Pan, Z.; Mao, Y. Non-hand-worn, load-free VR hand rehabilitation system assisted by deep learning based on ionic hydrogel. *Nano Res.* 2025, 18, 94907301.
- [3] Lu, L.; Hu, G.; Liu, J.; Yang, B. 5G NB-IoT System Integrated with High-Performance Fiber Sensor Inspired by Cirrus and Spider Structures. *Adv. Sci.* 2024, 11, e2309894.
- [4] Mengarelli, A.; Tigrini, A.; Scattolini, M.; Mobarak, R.; Burattini, L.; Fioretti, S.; Verdini, F. Myoelectric-Based Estimation of Vertical Ground Reaction Force During Unconstrained Walking by a Stacked One-Dimensional Convolutional Long Short-Term Memory Model. *Sensors* 2024, 24, 7768.
- [5] Tigrini, A.; Mobarak, R.; Mengarelli, A.; Khushaba, R.N.; Al-Timemy, A.H.; Verdini, F.; Gambi, E.; Fioretti, S.; Burattini, L. Phasor-Based Myoelectric Synergy Features: A Fast Hand-Crafted Feature Extraction Scheme for Boosting Performance in Gait Phase Recognition. *Sensors* 2024, 24, 5828.
- [6] Mobarak, R.; Tigrini, A.; Verdini, F.; Al-Timemy, A.H.; Fioretti, S.; Burattini, L.; Mengarelli, A. A Minimal and Multi-Source Recording Setup for Ankle Joint Kinematics Estimation During Walking Using Only Proximal Information From Lower Limb. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2024, 32, 812–821.
- [7] Vijayvargiya, A.; Singh, P.; Kumar, R.; Dey, N. Hardware Implementation for Lower Limb Surface EMG Measurement and Analysis Using Explainable AI for Activity Recognition. *IEEE Trans. Instrum. Meas.* 2022, 71, 1–9.
- [8] Moore, J.K.; Hnat, S.K.; van den Bogert, A.J. An elaborate data set on human gait and the effect of mechanical perturbations. *PeerJ* 2020, 3, e918.
- [9] Hu, B.; Rouse, E.; Hargrove, L. Benchmark Datasets for Bilateral Lower-Limb Neuromechanical Signals from Wearable Sensors During Unassisted Locomotion in Able-Bodied Individuals. *Front. Robot. AI* 2018, 5, 14.
- [10] Lencioni, T.; Carpinella, I.; Rabuffetti, M.; Marzegan, A.; Ferrarin, M. Human kinematic, kinetic and EMG data during different walking and stair ascending and descending tasks. *Sci. Data* 2019, 6.
- [11] Rodriguez-Martin, D.; Perez-Lopez, C.; Sama, A.; Catala, A.; Moreno Arostegui, J.M.; Cabestany, J.; Mestre, B.; Alcaine, S.; Prats, A.; Cruz Crespo, M.; et al. A Waist-Worn Inertial

Measurement Unit for Long-Term Monitoring of Parkinson's Disease Patients. *Sensors* 2017, 17, 827.

- [12] Suppa, A.; Kita, A.; Leodori, G.; Zampogna, A.; Nicolini, E.; Lorenzi, P.; Rao, R.; Irrera, F. 1-DOPA and Freezing of Gait in Parkinson's Disease: Objective Assessment through a Wearable Wireless System. *Front. Neurol.* 2017, 8, 406